CAR-TR-774
CS-TR-3484
May 1995

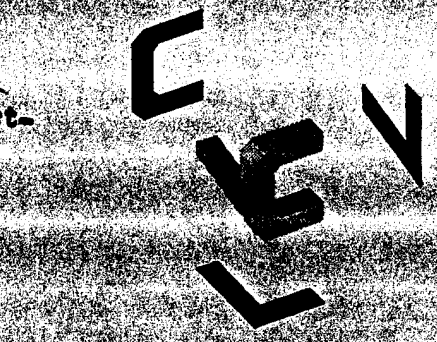
DACA76-92-C-0009
N00014-93-1-0257
IRI-90-57934

Perception of 3D Motion Through Patterns of Visual Motion

Cornelia Fermüller
Yiannis Aloimonos

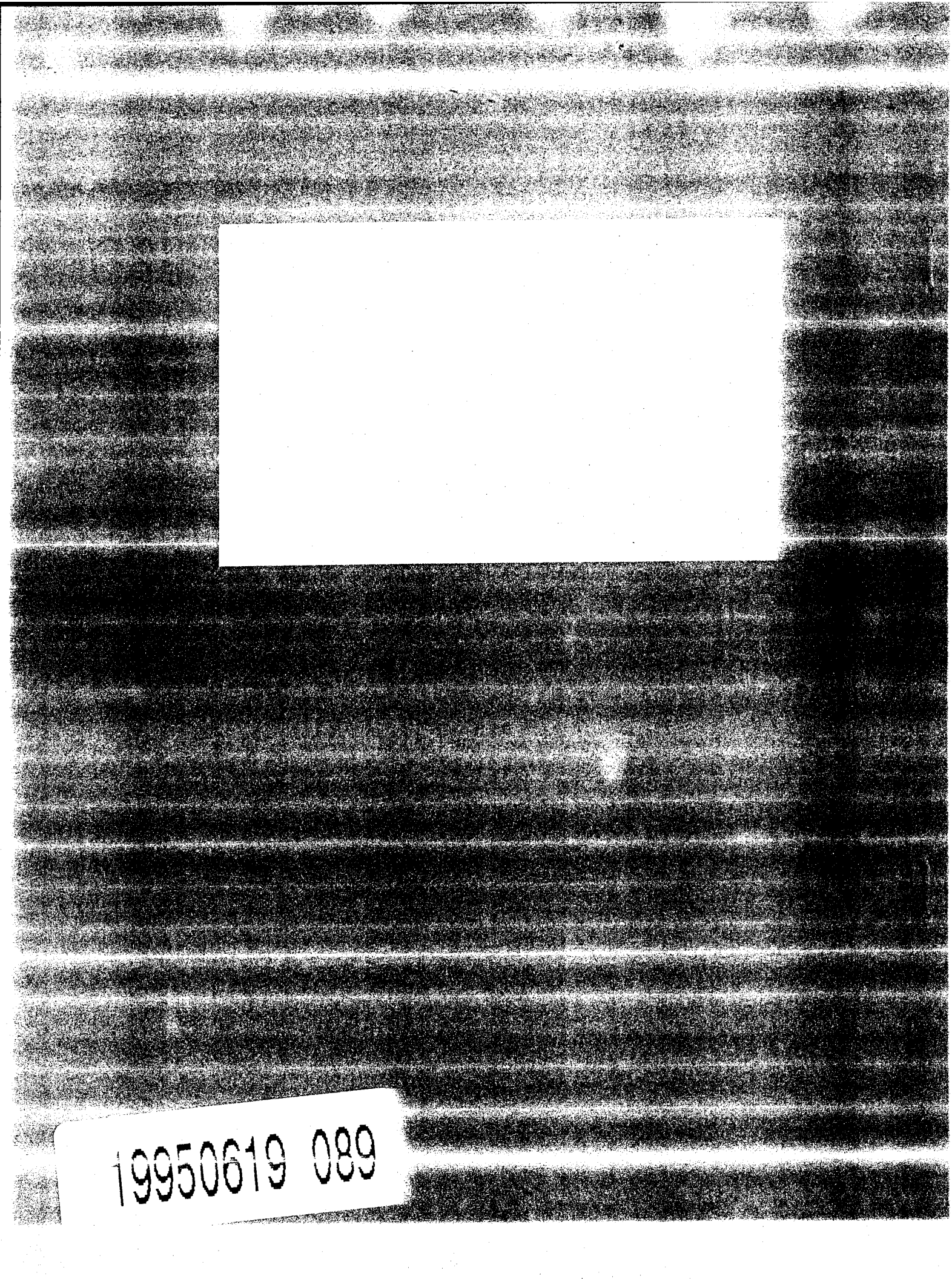
Computer Vision Laboratory
Center for Automation Research
Department of Computer Science and
Institute for Advanced Computer Studies
University of Maryland
College Park, MD 20742-3275

COMPUTER VISION LABORATORY

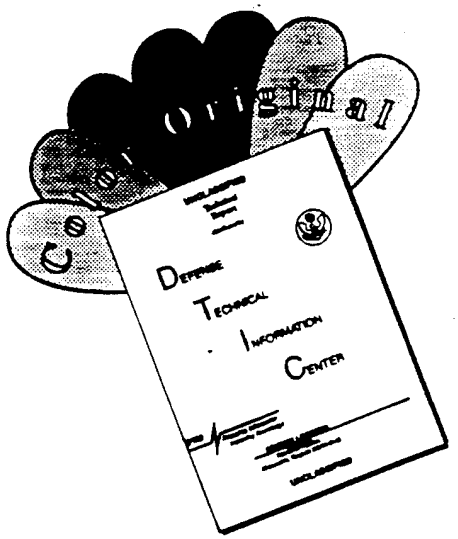


CENTER FOR AUTOMATION RESEARCH

UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND
20742-3275



DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF COLOR PAGES WHICH DO NOT REPRODUCE LEGIBLY ON BLACK AND WHITE MICROFICHE.

CAR-TR-774
CS-TR-3484
May 1995

DACA76-92-C-0009
N00014-93-1-0257
IRI-90-57934

Perception of 3D Motion Through Patterns of Visual Motion

Cornelia Fermüller
Yiannis Aloimonos

Computer Vision Laboratory
Center for Automation Research
Department of Computer Science and
Institute for Advanced Computer Studies
University of Maryland
College Park, MD 20742-3275

Abstract

Geometric considerations suggest that the problem of estimating a system's three-dimensional (3D) motion from a sequence of images, which has puzzled researchers in the fields of Computational Vision and Robotics as well as the Biological Sciences, can be addressed as a pattern recognition problem. Information for constructing the relevant patterns is found in spatial arrangements or gratings, that is, aggregations of orientations along which retinal motion information is estimated. The exact form of the gratings is defined by the shape of the retina or imaging surface; for a planar retina they are radial lines, concentric circles, as well as elliptic and hyperbolic curves, while for a spherical retina they become longitudinal and latitudinal circles for various axes. Considering retinal motion information computed normal to these gratings, patterns are found that have encoded in their shape and location on the retina subsets of the 3D motion parameters. The importance of these patterns is first that they depend only on the 3D motion and not on the scene in view, thus providing globally a separation of the effects of 3D motion and scene structure on the image motion, and second that they are founded upon easily derivable image measurements—they do not utilize exact retinal motion measurements such as optical flow, but only the sign of image motion along a set of directions defined by the gratings. The computational theory presented in this paper explains how the self-motion of a system can be estimated by locating these patterns. We also conjecture that this theory or variations of it might be implemented in nature and call for experiments in the neurosciences.

The support of the Advanced Research Projects Agency (ARPA Order No. 8459) and the U.S. Army Topographic Engineering Center under Contract DACA76-92-C-0009; the Office of Naval Research under Contract N00014-93-1-0257; and the National Science Foundation under Presidential Young Investigator Grant IRI-90-57934 is gratefully acknowledged, as is the help of Sandy German in preparing this paper.

To live is to move and to move is to live. All animals exist in space-time; they move in their environments and interact with it. Not surprisingly, to detect the sensory effects of movement is the first task of all sensory systems; and to reach an understanding of movement is a primary goal of all later perceptual analysis [3]. Although the organism as a whole might move in a nonrigid manner—head, arms, legs and wings undergo different motions—the eyes move rigidly, i.e., as a sum of instantaneous translation and rotation. The fundamental, abstract geometric concept used to describe the computational analysis of visual motion is that of the two-dimensional motion field: As a system moves in its environment, every point of the environment has a velocity vector with respect to the system. The projection of these three dimensional velocity vectors on the retina of the system's eye constitutes the so-called motion field. This field depends on the 3D motion and the structure of the scene in view. Considering a spherical eye moving with a translation $\vec{t}$, the motion field is along the great circles containing the vector $\vec{t}$ (Figure 1a), pointing away from the Focus of Expansion (FOE) and towards the Focus of Contraction (FOC). (The FOE and FOC are the points where $\vec{t}$ cuts the image sphere.) If the motion of the eye is a rotation of velocity $\vec{\omega}$ where the rotation axis cuts the sphere at points AOR (the Axis of Rotation point) and $-$AOR, the motion field is along the circles resulting from the intersection of the sphere with planes perpendicular to the rotation axis (Figure 1b). For general rigid motion the motion field on the sphere is the addition of a translational field and a rotational field (Figure 1c). In this case, the motion field does not have a simple structure and it becomes difficult to locate the points FOE and AOR, i.e., to solve the problem of egomotion using the two-dimensional motion field as input.

The problem is even more difficult, since what can be derived from the sequence of images sensed by the moving retina is not the exact projection of the 3D motion field but only information about the movement of light patterns. In the literature the exact movement of every point on the image is termed the optical flow field. In general, accurate values of the optical flow field are not computable. On the basis of local information only the component of the optical flow perpendicular to edges, the so-called normal flow, is well-defined (the aperture problem; see Figure 2). In many cases, it is possible to obtain additional flow information for areas (patches) in the image. Thus, the input that any system can use for

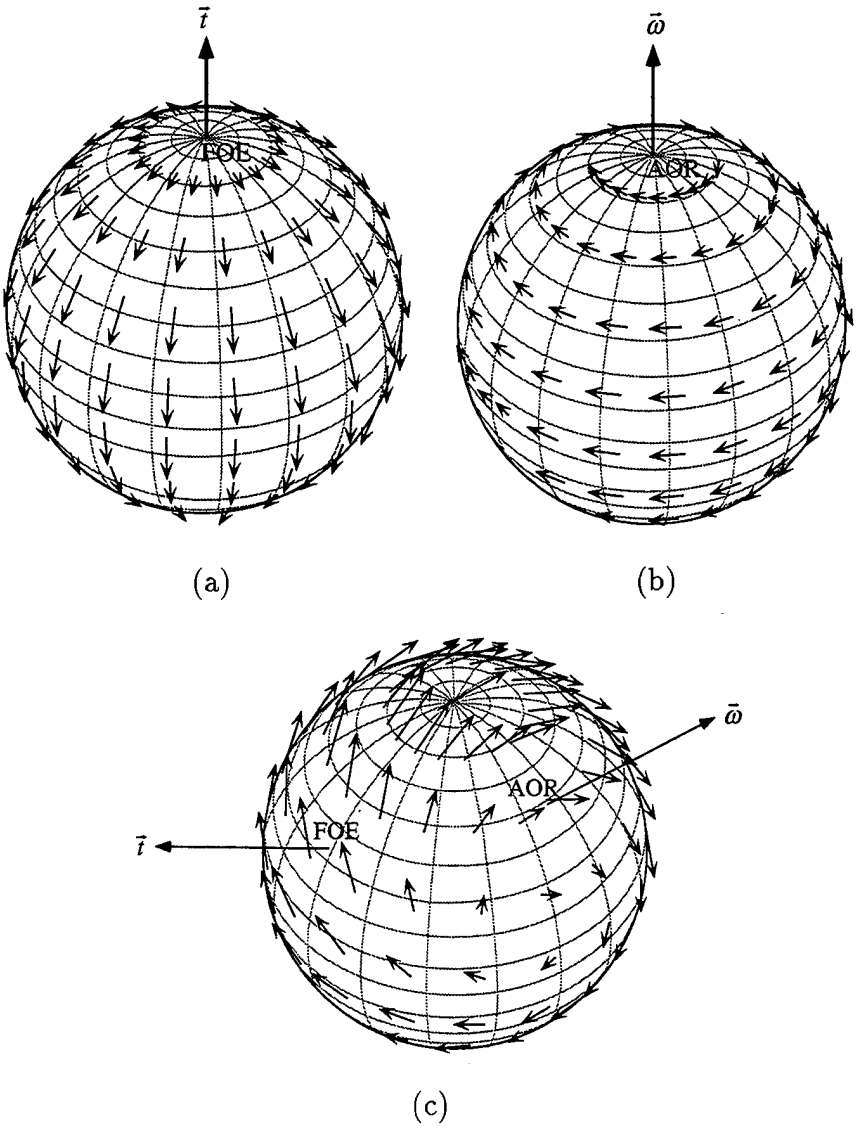


Figure 1: Motion fields on a spherical retina. The image $\vec{r}$ of a scene point with position vector $\vec{R}$ (with respect to an orthonormal coordinate system fixed to the center O of the sphere) is formed by perspective projection through center O. The sphere undergoes a rigid motion with translational velocity $\vec{t}$ and rotational velocity $\vec{\omega}$. (a) Translational motion field: at every point $\vec{r}$ the motion vector is $\frac{1}{|\vec{R}|}[(\vec{t}\cdot\vec{r})\vec{r}-\vec{t}]$, with $|\vec{R}|$ being the length of $\vec{R}$ and "$\cdot$" denoting inner vector product. Thus it is parallel to the great circle passing through the FOE and the FOC and its value is inversely proportional to the distance to the corresponding scene point. (b) Rotational motion field: at every point $\vec{r}$ the motion vector is $-\vec{\omega}\times\vec{r}$, where "$\times$" denotes outer vector product. Thus it is parallel to the circle passing through $\vec{r}$ perpendicular to $\vec{\omega}$, and it does not depend on the scene in view. (c) General rigid motion field: at every point $\vec{r}$ the motion vector is $\frac{1}{|\vec{R}|}[(\vec{t}\cdot\vec{r})\vec{r}-\vec{t}]-\vec{\omega}\times\vec{r}$.

further motion processing is partial optical flow information. The following analysis is based on a minimum amount of knowledge about image motion, namely the sign of the projection of optical flow along directions where it can be robustly computed. These measurements along a set of appropriately chosen orientations possess a rich global structure—they form simple patterns on the image surface whose location and form encodes the 3D motion parameters. A definition of the selected directions is given below.

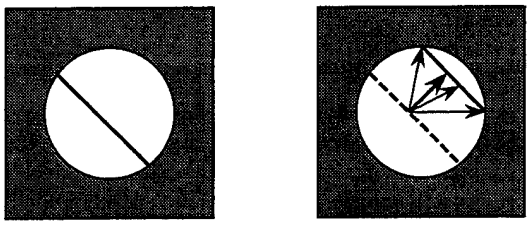
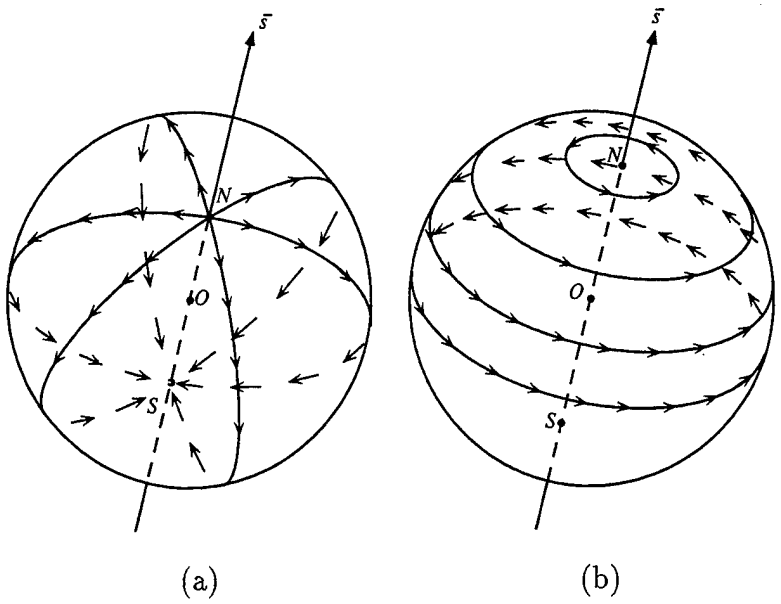


Figure 2: (a): Line feature observed through a small aperture at time t. (b): At time $t + \delta t$ the feature has moved to a new position. It is not possible to determine exactly where each point has moved to. From local measurements, only the flow component perpendicular to the line feature can be computed.

1 Selection of Flow Directions

Two classes of orientations are introduced which are defined with regard to an axis. Consider an axis $\vec{s}$ passing from the center of a spherical eye and cutting the sphere at points N and S. The unit vectors tangential to the great circles containing $\vec{s}$ define a direction for every point on the retina (Figure 3a). These directions are called $\vec{s}$-longitudinal. Similarly, the $\vec{s}$-latitudinal directions are defined as the unit vectors tangential to the circles resulting from the intersections of the sphere with planes perpendicular to the axis $\vec{s}$ (Figure 3b). At each point the $\vec{s}$-longitudinal and latitudinal vectors are perpendicular to each other.

Some properties of these directions will be of use later: Consider two axes $\vec{s}_1$ (N_1S_1) and $\vec{s}_2$ (N_2S_2). Each axis defines at every point a longitudinal and a latitudinal direction. The locus of points on the sphere where the $\vec{s}_1$-longitudinal directions are perpendicular to the $\vec{s}_2$-longitudinal directions (or where the $\vec{s}_1$ latitudinal directions are perpendicular to the $\vec{s}_2$ latitudinal directions) constitutes two quadratic curves whose geometry is explained in

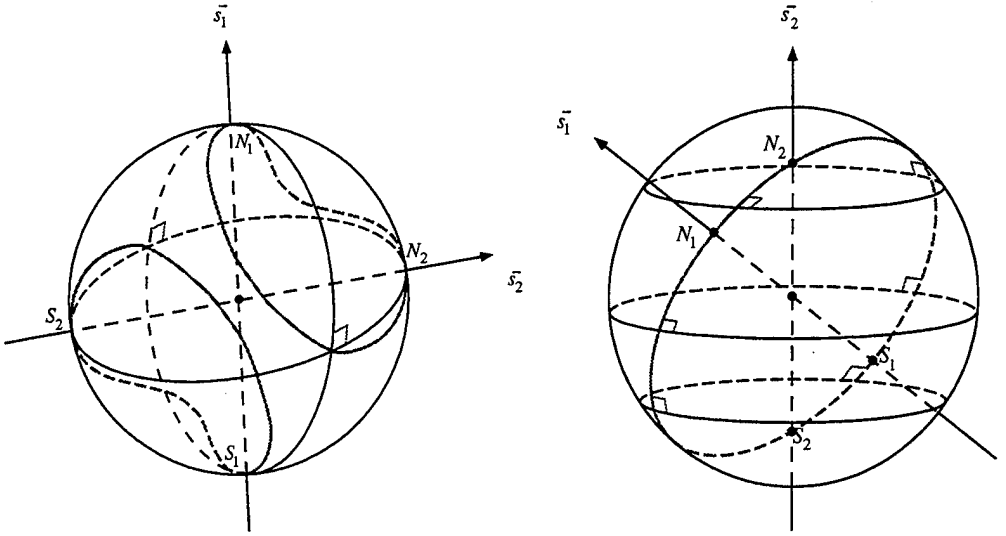


(a) (b)

Figure 3: (a) A longitudinal vector field defined by axis $\vec{s}$. At every point $\vec{r}$, a longitudinal vector has direction $(\vec{s} \cdot \vec{r})\vec{r} - \vec{s}$. (b) A latitudinal vector field defined by axis $\vec{s}$. At every point $\vec{r}$, a latitudinal vector has direction $-\vec{s} \times \vec{r}$.

Figure 4a. Similarly, the longitudinal directions of one axis and the latitudinal directions of the other axis are perpendicular to each other along the great circle defined by $\vec{s}_1$ and $\vec{s}_2$ (Figure 4b).

The structure of the projections of a rigid motion field on the $\vec{s}$-longitudinal and $\vec{s}$-latitudinal vectors will next be examined. More precisely, the signs of the projections of the motion field on the longitudinal and latitudinal vectors will be investigated, since this will be the information employed as input to the motion interpretation process. For this purpose it is necessary to agree upon a definition of the signs. $\vec{s}$ (NS)-longitudinal vectors are called positive $(+)$, if they point away from N, negative $(-)$ if they point away from S, and zero (0) otherwise. Similarly, $\vec{s}$-latitudinal vectors are referred to as positive $(+)$ if their direction is counterclockwise with respect to $\vec{s}$, negative $(-)$ if their direction is clockwise, and zero (0) otherwise (Figure 5).



(a) (b)

Figure 4: (a) On the sphere, the great circles containing $\vec{s}_1$ and $\vec{s}_2$ are perpendicular to each other on two closed second order curves, whose form depends on the angle between $\vec{s}_1$ and $\vec{s}_2$. These curves are defined as the set of points $\vec{r}$ for which $(\vec{s}_1 \times \vec{r}) \cdot (\vec{s}_2 \times \vec{r}) = 0$ or $(\vec{s}_1 \cdot \vec{r})(\vec{s}_2 \cdot \vec{r}) = \vec{s}_1 \cdot \vec{s}_2$. (b) The $\vec{s}_1$-longitudinal vectors are perpendicular to the $\vec{s}_2$-latitudinal vectors along the great circle through $\vec{s}_1$ and $\vec{s}_2$, defined as $(\vec{s}_1 \times \vec{s}_2) \cdot \vec{r} = 0$.

2 The Geometry of Image Motion Patterns

Since a rigid motion field is the addition of a translational and a rotational field, the cases of pure translation and pure rotation are first presented separately.

If the observer moves with a pure translation of velocity $\vec{t}$, the motion field on the sphere is along the direction of the $\vec{t}$-longitudinal vectors (Figure 1a). Projecting the translational motion field of Figure 1a on the $\vec{s}$-longitudinal vectors of Figure 3a, the resulting vectors will be either zero, positive or negative. The vectors will be zero on two curves as shown in Figure 4a (symmetric around the center of the sphere) whose shape depends on the angle between the vectors $\vec{t}$ and $\vec{s}$. The area inside the curves will contain negative vectors and the area outside the curves will contain positive vectors (Figure 6a).

If the observer moves purely rotationally with velocity $\vec{\omega}$, the motion field on the sphere is along the direction of the $\vec{\omega}$-latitudinal vectors (Figure 1b). Projecting the rotational motion field of Figure 1b on the $\vec{s}$ (NS)-longitudinal vectors of Figure 3a, the resulting vectors will

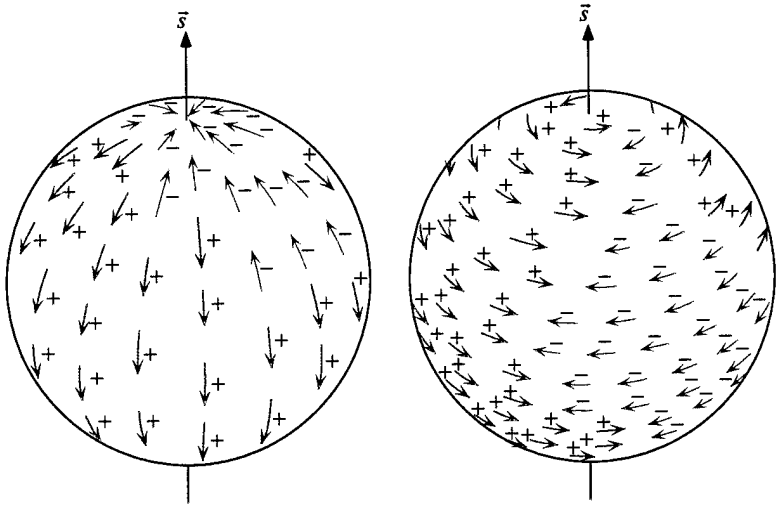
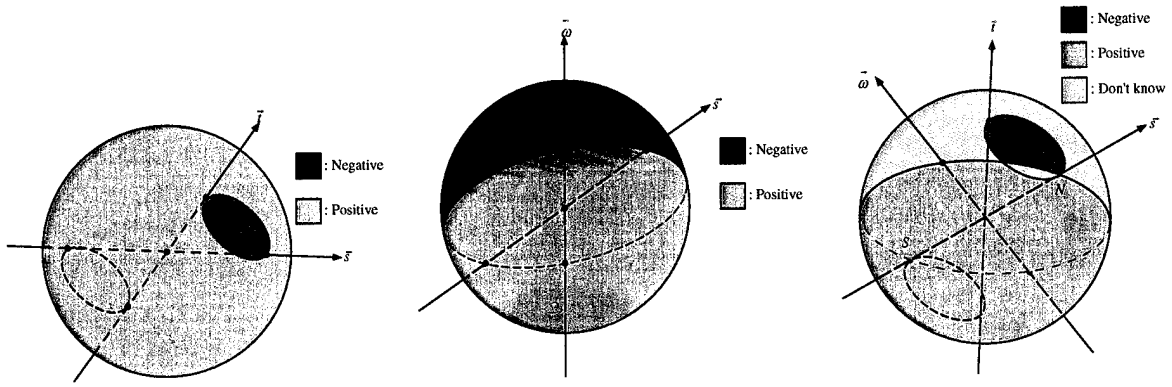


Figure 5: Positive and negative longitudinal (a) and latitudinal (b) image motion measurements. The input employed in the motion interpretation process is the sign of the image motion's value in the longitudinal and latitudinal directions.

be either zero, positive or negative. The projections will be zero on the great circle defined by $\vec{s}$ and $\vec{\omega}$, positive in one hemisphere and negative in the other (Figure 6b).

If the observer translates and rotates with velocities $\vec{t}$ and $\vec{\omega}$ the projection of the general motion field on any set of $\vec{s}$-longitudinal vectors can be classified for parts of the image. If at a longitudinal vector the projections of both the translational and the rotational vectors are positive, then the projection of the image motion (the sum of the translational and rotational vectors) will also be positive. Similarly, if the projections of both the translational and rotational vectors on a longitudinal vector are negative, the projection of the motion vector at this point will also be negative. In other words, if the values of Figures 6a and 6b are added, whenever positive and positive come together, the result will be positive, and whenever negative and negative come together, the result will be negative. However, whenever positive and negative come together, the result cannot be determined without knowledge of the environment. In such a case the sign of the projection of the rigid motion vector depends on the values of the translational and rotational vector components and thus on the lengths of the vectors $\vec{t}$ and $\vec{\omega}$ and the depth of the scene. (Actually, this "don't know" area also contains rich information regarding 3D motion and structure [4].)



(a) (b) (c)

Figure 6: $\vec{s}$-longitudinal pattern. (a) At every point $\vec{r}$ the projection of the translational motion vector on the $\vec{s}$-longitudinal vector is $\frac{1}{|\vec{R}|}[(\vec{t}\cdot\vec{r})\vec{r}-\vec{t}]\cdot[(\vec{s}\cdot\vec{r})\vec{r}-\vec{s}] = \frac{1}{|\vec{R}|}[\vec{s}\cdot\vec{t}-(\vec{s}\cdot\vec{r})(\vec{t}\cdot\vec{r})]$. It is zero on the curves $\vec{s}\cdot\vec{t} = (\vec{s}\cdot\vec{r})(\vec{t}\cdot\vec{r})$ (as shown in Figure 4a), negative inside the curves and positive outside the curves. (b) At every point $\vec{r}$ the projection of the rotational motion vector on the $\vec{s}$-longitudinal vector is $-(\vec{\omega}\times\vec{r})\cdot[(\vec{s}\cdot\vec{r})\vec{r}-\vec{s}] = (\vec{s}\times\vec{\omega})\cdot\vec{r}$. It is zero on the great circle $(\vec{s}\times\vec{\omega})\cdot\vec{r} = 0$ passing through $\vec{s}$ and $\vec{\omega}$, positive in one hemisphere and negative in the other. (c) A general rigid image motion defines a pattern along every $\vec{s}$-longitudinal vector field: an area of negative values, an area of positive values, and an area of values whose signs are unknown since they depend on the scene.

Thus, the distribution of the sign of image motion along any $\vec{s}$-longitudinal set of directions defines a pattern on the sphere. Considering a general rigid motion field due to translation $\vec{t}$ and rotation $\vec{\omega}$ on an $\vec{s}$ (NS)-longitudinal set of directions, a pattern like the one shown in Figure 6c is obtained, which consists of an area of strictly positive values, an area of strictly negative values, and an area in which the values cannot be determined without more information. The pattern is characterized by one great circle containing $\vec{\omega}$ and $\vec{s}$ and by two quadratic curves containing the points FOE, FOC, N and S.

It is worth stressing that the pattern of Figure 6c is independent of the scene in view and depends only on a subset of the 3D motion parameters. In particular, the great circle is defined by one rotational parameter and the quadratic curve by two translational parameters. Thus the pattern is of dimension three. Also, the pattern is different for a different choice of the vector $\vec{s}$. In summary, for a rigid motion $(\vec{t},\vec{\omega})$ with any axis $\vec{s}$ and a set of directions on the retina, an area of the imaging surface has been identified where the signs of the motion vectors along these directions does not depend on the scene in view!

Considering the projection of a rigid motion field on the $\vec{s}$ latitudinal directions (defined by the vector $\vec{s}\,(NS)$), another pattern (see Figure 7) is obtained which is dual to the one of Figure 6c. This time the translational latitudinal flow is separated into positive and negative by a great circle, and the rotational flow by two closed quadratic curves (as in Figure 3a) passing through the points AOR, $-$AOR, N and S.

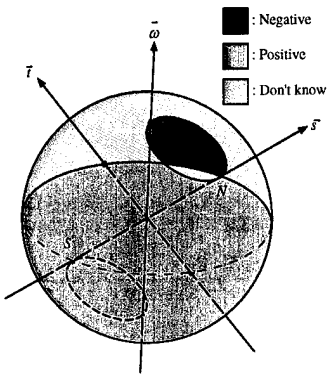


Figure 7: $\vec{s}$-latitudinal pattern. The translational flow is separated into positive and negative areas by a great circle and the rotational flow by two closed quadratic curves.

3 Egomotion Estimation Through Pattern Matching

The geometric analysis described above allows us to formulate the problem of egomotion estimation as a pattern recognition problem. Assume that the system has the capability of estimating the sign of the retinal motion along a set of directions defined by various $\vec{s}$-longitudinal or latitudinal fields. If the system can locate the patterns of Figures 6c and 7 in each longitudinal and latitudinal vector field, then it has effectively recognized the directions $\vec{t}$ and $\vec{\omega}$. The intersections of the quadratic curves of the patterns in Figure 6c and the great circles of the patterns in Figure 7 provide the points AOR and $-$AOR, and the intersections of the great circles of the patterns in Figure 6c and the quadratic curves of the patterns in Figure 7 provide the points FOE and FOC. Each single pattern provides only constraints on the locations of the FOE and the AOR, but a collection of patterns constrains these locations to small areas or even single points. It depends on the computational power of the system how much information will be available for pattern fitting and thus how accurately the FOE and the AOR can be localized. If the system is able to derive optical flow, then

it is able to estimate the sign of the projection of the flow along any direction and thus for every pattern at every point information is available. If, however, the system is less powerful and can only compute the motion in one direction (normal flow) or the sign of the motion in a few directions, then the solution proceeds exactly as before. The difference is that for each longitudinal or latitudinal set of directions, information (positive, negative or zero) is not available at every point of the sphere, and consequently the uncertainty may be larger and the FOE and AOR may be obtained only within bounds.

After the directions of $\vec{t}$ and $\vec{\omega}$ are estimated using the sign of the flow along various directions, the length of $\vec{\omega}$, i.e., the exact rotation, can easily be estimated using the values of the flow measurements. Also, after deriving the 3D motion from the information supplied by the patterns, the system could estimate optical flow for the purpose of deriving 3D scene structure and estimating the FOE and AOR more accurately. Usually, in a working system, information from other senses—such as inertial sensors—is utilized in addition.

4 Image Motion Patterns for a Planar Retina

For the case of a planar retina the latitudinal and longitudinal fields take a different form. There is a simple way of visualizing longitudinal and latitudinal fields on the sphere that carries through to the case of a planar retina. Given an axis $\vec{s}$, consider the family of cones with apex at the center of the sphere and axis $\vec{s}$. The intersection of these cones with the sphere provides circles lying on planes perpendicular to $\vec{s}$. The unit vectors perpendicular to the circles and tangential to the sphere form a longitudinal field. The intersection of the family of cones with a plane gives rise to a family of conic sections whose form (ellipses, hyperbolas, parabolas) depends on the angle between the axis $\vec{s}$ and the plane. The vectors perpendicular to these conics correspond to the longitudinal field (Figure 8a). On the sphere the latitudinal field is perpendicular to the great circles containing $\vec{s}$. These great circles when projected onto a plane become straight lines passing through a common point. The vectors perpendicular to these lines correspond to the latitudinal vector field (Figure 8b). The longitudinal fields on the plane are of a simple form when the chosen axis is parallel to an axis of a Cartesian 3D coordinate system attached to the observer; they are called α-fields

(axis parallel to the x-axis), β-fields (axis parallel to the y-axis) and γ-fields (axis parallel to the z or optical axis) (Figure 9). As before, the sign of the projection of the motion field on these particular fields possesses a rich structure as shown in Figure 10.

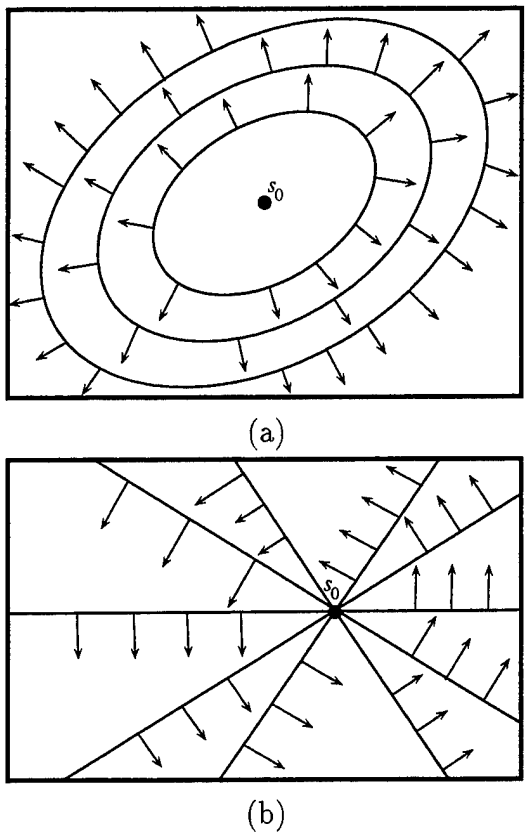


Figure 8: (a) In the plane the $\vec{s}$-longitudinal vectors become perpendicular to conic sections defined by a family of cones with axis parallel to $\vec{s}$. (b) The $\vec{s}$-latitudinal vectors become perpendicular to straight lines passing through the intersection s_0 of $\vec{s}$ with the plane.

Figures 11, 12 and 13 show results from experiments on synthetic spherical images and real planar images from an indoor and an outdoor scene.

5 Relationship to Other Computational Approaches

The pattern matching approach to egomotion estimation does not directly relate to traditional computational studies on the perception of 3D motion. Traditional studies, with a few exceptions [5, 6], addressed the problem in two steps. In the first step the optical flow field was estimated as an approximation to the motion field. In the second step, the 3D motion

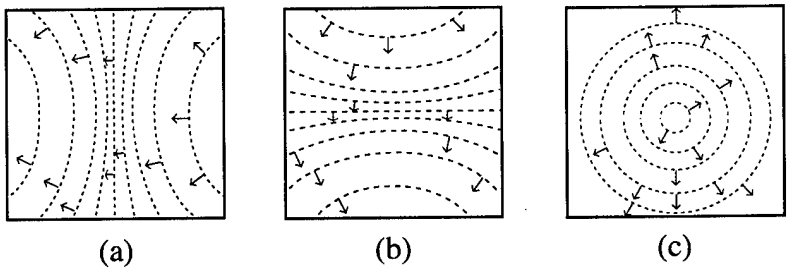


Figure 9: Positive α, β and γ vectors. If the $\vec{s}$-axis is the x, y or z-axis the corresponding longitudinal vector fields are vectors perpendicular to: (a) and (b) hyperbolas whose axes coincide with the vertical and horizontal axes on the image plane, or (c) concentric circles with their center at the origin of the image.

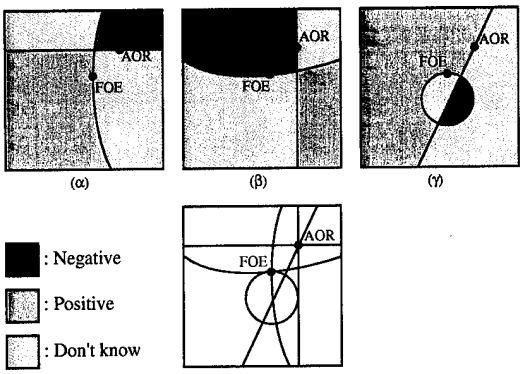


Figure 10: α, β and γ patterns. (α) Image motion measurements along the α-vectors form patterns of positive and negative values which are defined by a horizontal straight line and a hyperbola. (β) β-patterns are defined by a vertical straight line and a hyperbola and (γ) γ-patterns are defined by a straight line through the image center and a circle passing through the center of the image. The intersection of the straight lines gives the AOR and the intersection of the conics gives the FOE.

was estimated through a local decomposition of the optical flow field [7–11]. In the scheme described here, the retinal motion information utilized consists of the signs of the optical flow along certain directions. In other words, for a vector $\vec{v}$ on the image, the information needed is whether the flow along the line defined by $\vec{v}$ has the sign of $\vec{v}$ or $-\vec{v}$. This is a robust qualitative property of the optical flow, and as demonstrated here, it is sufficient for the task of egomotion perception. In the literature, it has been argued [12, 13] that qualitative estimates of optical flow are often sufficient for many tasks; for instance, for the task of detecting a potential crash [14], not even a precise measurement of the normal component of the flow may be necessary. As suggested in [15], "it is sufficient that the image motion estimate be qualitatively consistent with the perspective 2D projection of the 'true' 3D ve-

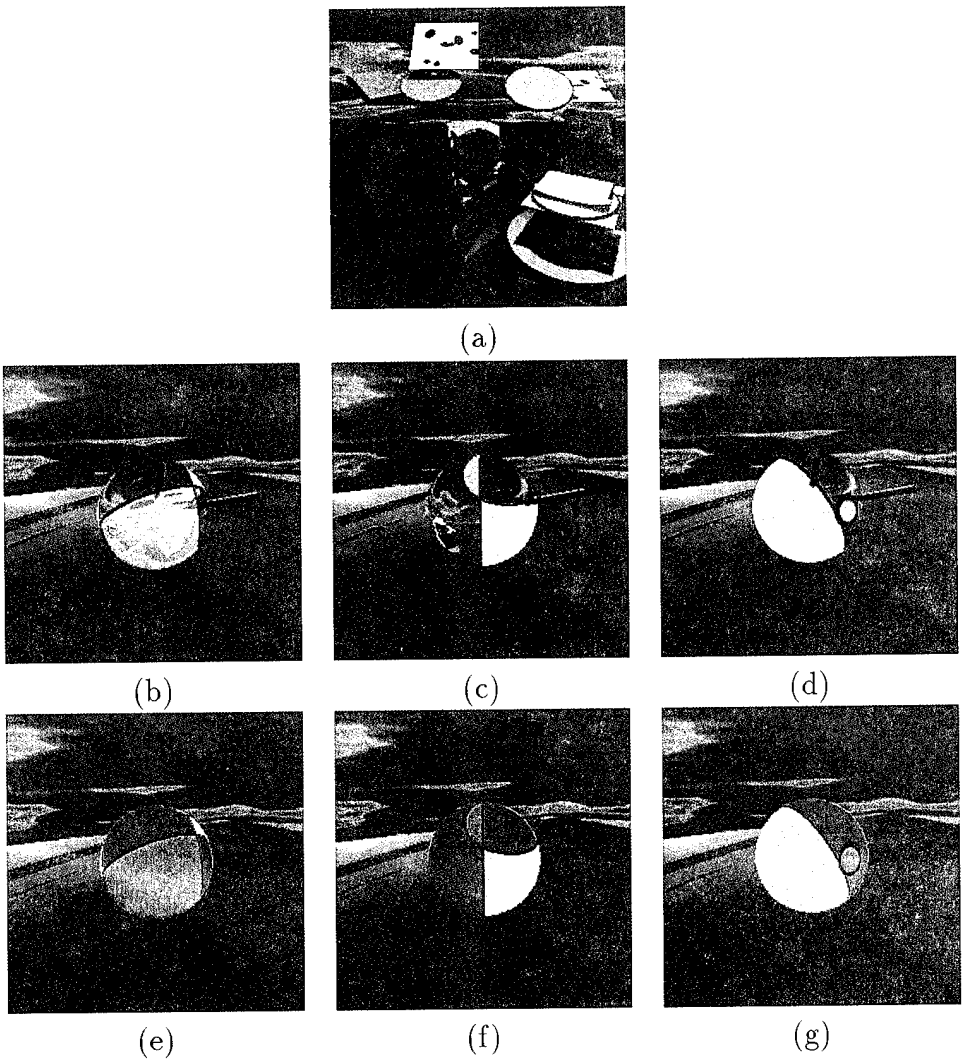


Figure 11: (a) Using a graphics package the image of a synthetic environment (a glass table set up in front of a window—notice the reflection on the glass and the sky background) including a sphere (a blue ball in front of a cheese) has been created. From the resulting images the normal flow has been computed. (b), (c) and (d) show, overlaid on the spherical image, the positive and negative areas for three different patterns; red corresponds to positive and green to negative. The motion is graphically rendered through its translation axis (black axis) and its rotation axis (grey axis). The three axes $\vec{s}$ chosen for the patterns are the x-, y- and z-axes. (e), (f) and (g) show for the same configuration the spherical retinae containing only the patterns (pink denoting positive areas, turquoise negative areas, and blue "don't-care" areas).

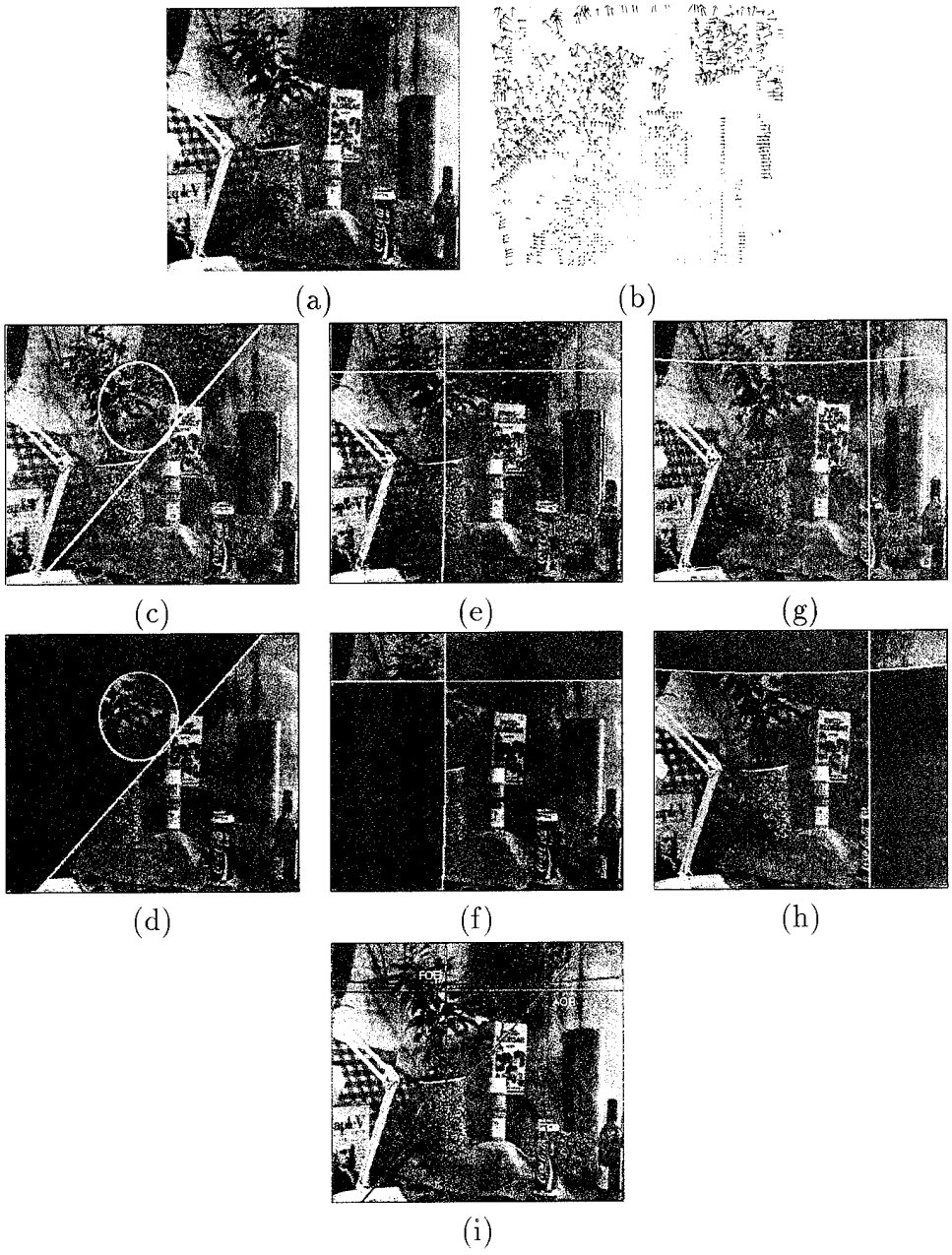


Figure 12: For a sequence of images taken by a sensor moving rigidly, the normal flow field was estimated. The image size is 574×652, the focal length in the x-direction is 1163 pixels, and the focal length in the y-direction is 1362 pixels. The center of the image is $(332, 305)$ (measuring from the bottom left corner), the FOE is $(255, 129)$, and the AOR is $(496, 160)$. (a) shows the first frame of the sequence and (b) shows the normal flow field corresponding to the first and second frame. (c) shows the positive and negative γ-vectors found from the normal flow field (blue denotes negative and red denotes positive) and (d) shows the fitting of the γ-pattern to the γ-vectors in the final stage, after all patterns have been computed. Similarly, (e) and (g) show the positive and negative α- and β-vectors, and (f) and (h) the fitting of the α-pattern to the α-vectors and the β-pattern to the β-vectors in the final stage. (i) shows the curves (red) and the straight lines (blue) of the α-, β-, and γ-patterns superimposed on the image. The intersection of the second order curves provides the FOE and the intersection of the straight lines gives the AOR.

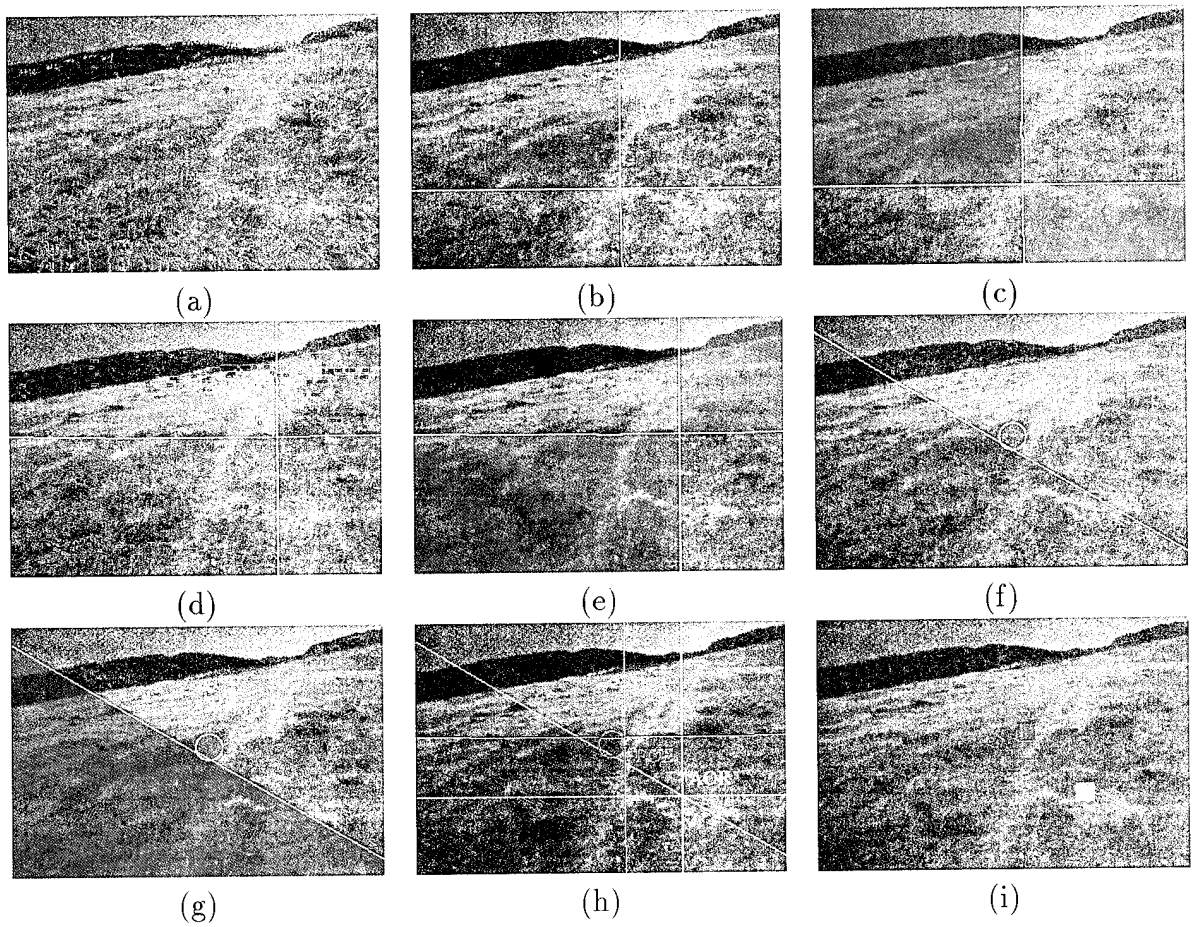


Figure 13: A camera mounted on the Unmanned Ground Vehicle, developed by Martin Marietta Corp. under a contract with the U.S. Government, captured a sequence of images as the vehicle moved along rough terrain in the countryside, thus undergoing continuously changing rigid motion. (a) shows one frame of the sequence with the normal flow field overlaid in red. (b), (d) and (f) show the positive α-, β- and γ-vectors and (c), (e) and (g) show fitted α-, β- and γ-patterns in the final stage after all the patterns have been computed. (h) shows superimposed on the image the boundaries of the patterns whose intersections provide the FOE and the AOR. (i) Because measurements are not everywhere available (strong spatial gradients appear sparse), a set of patterns can be fitted with accuracy above the threshold of 97% (where accuracy is defined as the ratio of the number of successfully fitted pixels over the total number of pixels in the pattern), resulting in solutions for the FOE and the AOR lying within two bounded areas (red: FOE; green: AOR).

locity field. Even estimates that don't correspond to image velocity, like the ones derived by Reichardt's correlation model or equivalent energy models [16–19], may be acceptable for several visual tasks if the estimates are consistent over the visual field." The pattern-based approach to the problem of egomotion estimation proves the feasibility of such ideas about qualitative visual motion analysis.

6 The Motion Pathway in Primates

The computational framework described here is consistent with findings in neurobiology regarding the structure and functional properties of neurons in the visual motion pathway (Figure 14) [20]. According to Movshon (and many others), in the early stages, from the retinal Pa ganglion cells through the magnocellular LGN cells to layer 4Ca of V1 the cells appear functionally homogeneous and respond almost equally well to the movement of a bar (moving perpendicularly to its direction) in any direction (Figure 14a). Within layer 4C of V1 an onset of orientation selectivity is observed. The receptive fields of the neurons here are divided into separate excitatory and inhibitory regions. The regions are arranged in parallel stripes and this arrangement provides the neurons with a preference for a particular orientation of a bar target (which is displayed in the polar diagram) (Figure 14b). In layer 4B of V1 another major transformation takes place with the appearance of directional selectivity. The receptive fields here are relatively large and they seem to be excited everywhere by light or dark targets. In addition, these neurons respond better or solely to one direction of motion of an optimally oriented bar target, and less or not at all to the other (Figure 14c). In MT neurons have considerably large receptive fields and in general the precision of the selectivity for direction of motion that the neurons exhibit is typically less than in V1 (Figure 14d). In MST the size of the receptive fields of neurons becomes even larger, ranging from 30 degrees to 100 degrees, each responding to particular 3D motion configurations [1, 2, 21–24].

One can easily envision an architecture that, using neurons with the properties listed above, implements a global decomposition of the motion field using the signs of the motion vectors along appropriately chosen directions. Neurons of the first kind could be involved in the estimation of local retinal motion information; they could be thought of as computing

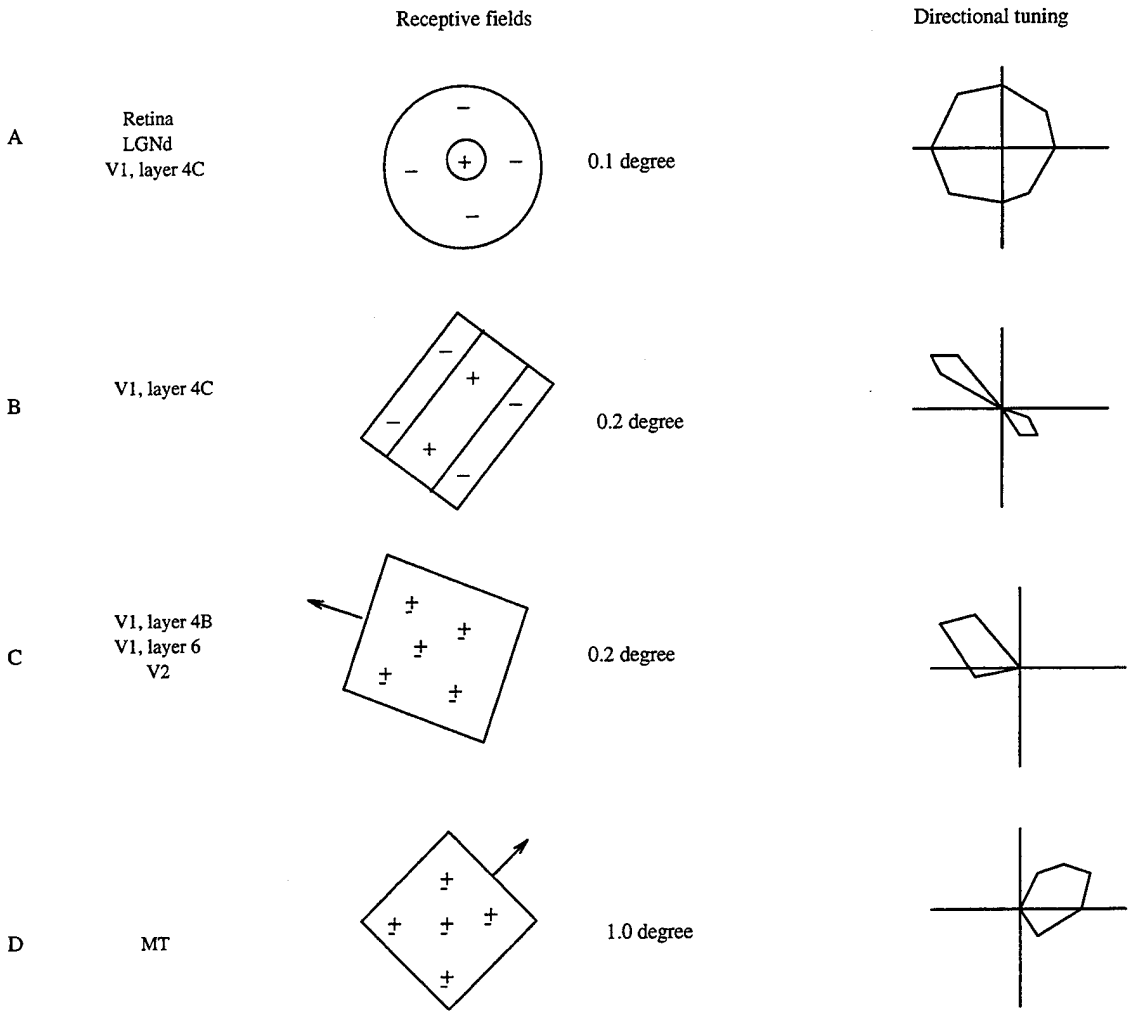


Figure 14: The spatial structure of visual receptive fields and their directional selectivity at different levels of the motion pathway (from [20]). The spatial scales of the receptive fields (0.1 degree, etc.) listed here are for neurons at the center of gaze; in the periphery these dimensions would be larger. The polar diagrams illustrate responses to variation in the direction of a bar target oriented at right angles to its direction of motion. The angular coordinate in the polar diagram indicates the direction of motion and the radial coordinate the magnitude of the response.

whether the projection of retinal motion along some direction is positive or negative. Neurons of the second kind could be involved in the selection of local vectors in particular directions as parts of the various different patterns discussed in this paper, while neurons of the third kind could be involved in computing the sign (positive or negative) of pattern vectors for areas in the image; i.e., they might compute, for image patches of different sizes, whether the flow in certain directions is positive or negative. Finally, neurons of the last kind (MT and MST) could be the ones that piece together the parts of the patterns found already into global

patterns that are matched with prestored global patterns. Matches provide information about egomotion and mismatches provide information about independent motion.

Results from the cognitive sciences have suggested that a large part of visual perception may be realized as pattern matching. Also, most of the early work in computational vision was concerned with the development of general pattern matching techniques, without paying much attention to the nature of visual patterns. The body of work conducted in computational modeling and perceptual disciplines in the last decades (see for example [1, 2, 21–27]) now provides the prerequisites for addressing the difficult question of what patterns are relevant to particular visual tasks.

The pattern matching techniques described here for handling visual motion were based on purely theoretical considerations. It remains to be shown if such schemes or variants of them are actually implemented in biological organisms, whether these are insects possessing spherical-like compound eyes [28, 29] or primates with elaborate and sophisticated motion processing capabilities. Recent findings provide a good motivation: in area V4 of the macaque monkey neurons have been found [30] whose function has been linked to shape processing. These neurons respond to gratings of the same nature as those utilized in the patterns of this paper!

References

[1] C.J. Duffy and R.H. Wurtz. Sensitivity of MST neurons to optical flow stimuli I: A continuum of response selectivity to large field stimuli. *Journal of Neurophysiology*, 65:1329–1345, 1991.

[2] K. Tanaka and H.A. Saito. Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells illustrated in the dorsal part of the Medial Superior Temporal area of the macaque monkey. *Journal of Neurophysiology*, 62:626–641, 1989.

[3] Y. Aloimonos and A. Rosenfeld. Computer vision. *Science*, 253:1181–1324, 1991.

[4] C. Fermüller and Y. Aloimonos. On the geometry of visual correspondence. Technical Report CAR-TR-732, Center for Automation Research, University of Maryland, 1994.

[5] J. Aloimonos and C.M. Brown. Direct processing of curvilinear sensor motion from a sequence of perspective images. In *Proc. Workshop on Computer Vision: Representation and Control*, pages 72–77, 1984.

[6] B.K.P. Horn and E.J. Weldon. Computationally efficient methods for recovering translational motion. In *Proc. International Conference on Computer Vision*, pages 2–11, 1987.

[7] H.C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

[8] H.C. Longuet-Higgins and K. Prazdny. The interpretation of a moving retinal image. *Proc. Royal Soc. London B*, 208:385–397, 1984.

[9] J.J. Koenderink. Optic flow. *Vision Research*, 26:161–180, 1986.

[10] R.Y. Tsai and T.S. Huang. Uniqueness and estimation of three dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:13–27, 1984.

[11] S. Ullman. *The Interpretation of Visual Motion.* MIT Press, Cambridge, MA, 1979.

[12] T. Poggio, A. Verri, and V. Torre. Green theorems and qualitative properties of the optical flow. A.I. Memo 1289, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, 1991.

[13] A. Verri and T. Poggio. Motion field and optical flow: Qualitative properties. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:490–498, 1989.

[14] R.C. Nelson and J.Y. Aloimonos. Obstacle avoidance using flow field divergence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:1102–1106, 1989.

[15] N. Ancona and T. Poggio. Optical flow from 1D correlation: Application to a simple time-to-crash detector. *International Journal of Computer Vision*, in press. Special Issue on Qualitative Vision, Y. Aloimonos, Ed.

[16] W. Reichardt. Autocorrelation, a principle for evaluation of sensory information by the central nervous system. In W.A. Rosenblith, Ed., *Principles of Sensory Communication*, pages 303–317. Wiley, New York, 1961.

[17] W. Reichardt. Evaluation of optical motion information by movement detectors. *Journal of Computational Physics*, 161:533–547, 1987.

[18] T. Poggio and W. Reichardt. Considerations on models of movement detection. *Kybernetik*, 13:223–227, 1973.

[19] J.P.H. van Santen and G. Sperling. Temporal covariance model of human motion perception. *Journal of the Optical Society of America A*, 1:451–473, 1984.

[20] A. Movshon. Visual processing of moving images. In H. Barlow, C. Blakemore, and M. Weston-Smith, Eds., *Images and Understanding*, pages 122–137. Cambridge University Press, 1990.

[21] R.A. Andersen, R.J. Snowden, S. Treue, and M. Graziano. Hierarchical processing of motion in the visual cortex of monkey. In *The Brain* (Cold Spring Harbor Symposium on Quantitative Biology **55**), pages 741–748. Cold Spring Harbor Laboratory Press, 1990.

[22] C.J. Duffy and R.H. Wurtz. Sensitivity of MST neurons to optical flow stimuli ii: Mechanisms of response selectivity revealed by small field stimuli. *Journal of Neurophysiology*, 65:1346–1359, 1991.

[23] L. Lagae. *A neurophysiological Study of Optic Flow Analysis in the Monkey Brain.* PhD thesis, Leuven, Belgium, 1991.

[24] G.A. Orban, L. Lagae, A. Verri, S. Raiguel, D. Xiao, H. Maes, and V. Torre. First order analysis of optical flow in monkey brain. *Proc. Natl. Acad. Sci. U.S.A.*, 1992, in press.

[25] N.K. Logothetis, J. Pauls, and T. Poggio. Viewer-centered object recognition in monkeys. A.I. Memo 1473, Artificial Intelligence Laboratory and Center for Biological and Computational Learning, Massachusetts Institute of Technology, Cambridge, MA, April 1994.

[26] B.K.P. Horn. *Robot Vision.* McGraw Hill, New York, 1986.

[27] J. Aloimonos and D. Shulman. *Integration of Visual Modules: An Extension of the Marr Paradigm.* Academic Press, Boston, 1993.

[28] G.A. Horridge. The evolution of visual processing and the construction of seeing systems. *Proceedings of the Royal Society, London B*, 230:279–292, 1987.

[29] R. Wehner. Homing in arthropods. In F. Papi, Ed., *Animal Homing*, pages 45–144. Chapman and Hall, London, 1992.

[30] J.L. Gallant, J. Braun, and D.C. Van Essen. Selectivity for polar, hyperbolic, and cartesian gratings in macaque visual cortex. *Science*, 259:100–103, 1993.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May 1995	Technical Report

4. TITLE AND SUBTITLE
Perception of 3D Motion Through Patterns of Visual Motion

5. FUNDING NUMBERS
DACA76-92-C-0009
N00014-93-1-0257
IRI-90-57934

6. AUTHOR(S)
Cornelia Fermüller
Yiannis Aloimonos

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Computer Vision Laboratory
Center for Automation Research
University of Maryland
College Park, MD 20742-3275

8. PERFORMING ORGANIZATION REPORT NUMBER
CAR-TR-774
CS-TR-3484

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Advanced Research Projects Agency, 3701 N. Fairfax Dr., Arlington, VA 22203-1714
U.S. Army Topographic Engineering Center, 7701 Telegraph Road, Bldg. #2592, Alexandria, VA 22310-3864
Office of Naval Research, 800 N. Quincy St., Arlington, VA 22217-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release.
Distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Geometric considerations suggest that the problem of estimating a system's three-dimensional (3D) motion from a sequence of images, which has puzzled researchers in the fields of Computational Vision and Robotics as well as the Biological Sciences, can be addressed as a pattern recognition problem. Information for constructing the relevant patterns is found in spatial arrangements or gratings, that is, aggregations of orientations along which retinal motion information is estimated. The exact form of the gratings is defined by the shape of the retina or imaging surface; for a planar retina they are radial lines, concentric circles, as well as elliptic and hyperbolic curves, while for a spherical retina they become longitudinal and latitudinal circles for various axes. Considering retinal motion information computed normal to these gratings, patterns are found that have encoded in their shape and location on the retina subsets of the 3D motion parameters. The importance of these patterns is first that they depend only on the 3D motion and not on the scene in view, and second that they utilize only the sign of image motion along a set of directions defined by the gratings.

14. SUBJECT TERMS
Motion perception, primate vision, visual motion

15. NUMBER OF PAGES
24

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements.***

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	**PR**	- Project
G	- Grant	**TA**	- Task
PE	- Program Element	**WU**	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

- **DOD** - See DoDD 5230.24, "Distribution Statements on Technical Documents."
- **DOE** - See authorities.
- **NASA** - See Handbook NHB 2200.2.
- **NTIS** - Leave blank.

Block 12b. Distribution Code.

- **DOD** - Leave blank.
- **DOE** - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
- **NASA** - Leave blank.
- **NTIS** - Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.